Complexity and approximation of the smallest k-enclosing ball problem

Vladimir Shenmaier[1]

Abstract. Given an n-point set in Euclidean space $\mathbb{R}^d$ and an integer k, consider the problem of finding the smallest ball enclosing at least k of the points. In the case of a fixed dimension the problem is polynomial-time solvable but in the general case, when d is not fixed, the complexity status of the problem was not yet known. We prove that the problem is strongly NP-hard and describe an idea of PTAS.

The Smallest k-Enclosing Ball problem is considered:

Problem S*k*-EB Given a set X of n points in Euclidean space $\mathbb{R}^d$ and an integer k. Find the smallest ball enclosing at least k of the points.

The problem has a lot of interesting interpretations of life, due to simplicity of the formulation. One of them is in the area of military affairs: given coordinates of n purposes, hit k of them in one gulp with a minimum charge.

Related work

The earliest reference of a special case of this problem occurs in the middle of the 19th century [9]. In the case of a fixed dimension, particularly Euclidean plane (the most studied case), the problem is polynomially solvable [2]. However, the running time of the best known algorithms depends exponentially on dimension [4,6]. In the general case, when d is not fixed (belongs to a problem instance), the complexity status of the problem was not yet known. Agarwal *et al.* [1] study a very related problem "smallest enclosing ball with outliers" and present an approximation scheme (PTAS) for high dimensions based on coresets.

[1] Sobolev Institute of Mathematics, Novosibirsk, Russia. Email: shenmaier@mail.ru. This research is supported by RFBR (projects 12-01-00093, 12-01-33028 and 13-01-91370ST), Presidium RAS (project 227) and IM SBRAS (project 7B).

J. Nešetřil et al. (eds.), *The Seventh European Conference on Combinatorics, Graph Theory and Applications*

Our results

We prove that the problem is strongly NP-hard and unless P = NP there is no fully polynomial-time approximation scheme (FPTAS) (that does not follow from the strong NP-hardness when solution values are not integer). Also we describe a straightforward idea of PTAS that computes a $(1+\varepsilon)$-approximation in $O(n^{1/\varepsilon^2+1}d)$ time for any $\varepsilon > 0$.

1 Hardness results

We formulate a special case of Sk-EB in the form of the problem of verification of properties on a set of Boolean vectors:

Special Case Given a set X of n Boolean points, $X \subseteq \{0, 1\}^d$, an integer $k \in [1, n]$, and a real value $R > 0$. Determine wether there is a Euclidean ball of radius R enclosing at least k of the points.

Using ideas from the work of [7], we give a reduction to this special case from the following strongly NP-complete problem [8]:

Clique in Regular Graph Given a regular graph G and an integer k. Determine wether there is a complete k-vertex subgraph of G.

Theorem 1.1. *Clique in Regular Graph can be reduced to Special Case in polynomial time.*

Reduction. Let G be a regular graph on n vertices and m edges, and Δ be the degree of the vertices of G. Define a set X as the set of m-dimensional rows of the incidence matrix of G. Observe that any two points x and y of X have a distance $\sqrt{2\Delta - 2}$ if the corresponding vertices are adjacent, and $\sqrt{2\Delta}$ otherwise.

For any k-point set K in $\mathbb{R}^d$, define an average of K as $\overline{c}(K) = \sum_{x\in K} x \big/ k$, and for any $y \in \mathbb{R}^d$, define a value $f(y, K) = \sum_{x\in K} \|x - y\|^2$.

Lemma 1.2. $f(y, K) = f(\overline{c}(K), K) + k\,\|y - \overline{c}(K)\|^2$.

Proof. Indeed, $f(y, K) = \sum_{x\in K} \|x - \overline{c}(K)\|^2 - \sum_{x\in K} 2\,\langle x - \overline{c}(K), y - \overline{c}(K)\rangle + \sum_{x\in K} \|y - \overline{c}(K)\|^2$, where $\langle .\,, .\rangle$ is scalar product of vectors. The first term in this expression is equal to $f(\overline{c}(K), K)$, the latest is equal to $k\,\|y - \overline{c}(K)\|^2$, and the second is equal to zero, since $\sum\limits_{x\in K} x = k\,\overline{c}(K)$. □

For any k-point set K in $\mathbb{R}^d$, define a value $g(K) = \sum_{x\in K} \sum_{y\in K} \|x - y\|^2$.

Lemma 1.3. $g(K) = 2k\, f(\overline{c}(K), K)$.

Proof. This follows from Lemma 1.2. □

Lemma 1.4. *Let R be the radius of the smallest ball enclosing at least k of the points of X, and $A = (1 - 1/k)(\Delta - 1)$. Then $R^2 \le A$, if there is a k-clique in the graph G, and $R^2 \ge A + 2/k^2$ otherwise.*

Proof. Suppose that a k-clique exists. Consider the point $\overline{c}(K)$, where K is the set of points corresponding to the vertices of the clique. Using Lemma 1.3, we have $f(\overline{c}(K), K) = g(K)/2k = (k^2 - k)(2\Delta - 2)/2k = k\,A$. By symmetry, distances from $\overline{c}(K)$ to all the points of K are the same. It follows that the squares of these distances are equal to the value $f(\overline{c}(K), K)/k = A$. Thus, the ball of radius $\sqrt{A}$ centered at the point $\overline{c}(K)$ covers all the points of K.

Suppose that there is no k-clique. Let c be the center of the smallest ball enclosing k points of X, and K be the k-point subset it covers. Then R is equal to the maximal distance between c and the points of K. Since the maximum of squares of distances is at least its average, we have $R^2 \ge f(c, K)/k$. On the other hand, from Lemma 1.2 it follows that $f(c, K) \ge f(\overline{c}(K), K)$. Therefore, $R^2 \ge f(\overline{c}(K), K)/k = g(K)/2k^2$. But any of $k^2 - k$ summands in the definition of $g(K)$ corresponding to the pairs of distinct points is either 2Δ or $2\Delta - 2$. And, by the assumption of nonexistence of k-clique, at least two of them equals 2Δ. Therefore, $g(K) \ge (k^2 - k)(2\Delta - 2) + 4$, and then $R^2 \ge (1 - 1/k)(\Delta - 1) + 2/k^2 = A + 2/k^2$. □

Proof of Theorem 1.1. Lemma 1.4 implies that existence of a k-clique in graph G is equivalent to existence of a ball of radius $\sqrt{A}$ enclosing k points of X. This completes the proof of Theorem 1.1. □

Corollary 1.5. *The Smallest k-Enclosing Ball problem is strongly* NP-*hard.*

In the case of optimization problems with integer-valued solutions the strong NP-hardness implies that there is no fully polynomial-time approximation scheme (FPTAS) [5, 10]. Unfortunately, the radius of the smallest k-enclosing ball is not integer in general, and so that fact needs proof.

Theorem 1.6. *For the Smallest k-Enclosing Ball problem, there is no fully polynomial-time approximation scheme (FPTAS) unless* P = NP.

Proof. By Lemma 1.4, if there is a k-clique in a graph G then the radius of the smallest ball enclosing k of the points of X is bounded by $\sqrt{A}$, and otherwise it is at least $\sqrt{A + 2/k^2} > \sqrt{A}\,(1 + 1/2Ak^2) > \sqrt{A}\,(1 + 1/2n^3)$. It follows that there is no polynomial-time algorithm that computes a $(1+1/2n^3)$-approximation unless $\mathsf{P} = \mathsf{NP}$. On the other hand, for an arbitrary polynomial $p(n)$, any FPTAS allows to compute a $(1 + 1/p(n))$-approximation in polynomial time. Therefore, existence of FPTAS is impossible. □

2 Approximation scheme

We describe an idea of a polynomial-time approximation scheme (PTAS) for Sk-EB based on the simple gradient descent type algorithm [3] for the Small Enclosing Ball problem:

Problem SEB Given a set X of n points in Euclidean space $\mathbb{R}^d$. Find the smallest ball enclosing all the points.

Observe that this is a special case of Sk-EB: $k = n$. The following approximation algorithm is considered in [3]:

Algorithm for SEB Let i be an arbitrary positive integer.
`Step 1:` Choose any point $c_1 \in K$.
`Step` j, $j = 2, 3, \ldots, i$: Take a point $p_j \in K$, which is furthest away from c_{j-1}, and define $c_j = c_{j-1} + (p_j - c_{j-1})/j$.
`Output:` The ball of radius $R(c_i, K) = \max_{x \in K} \|x - c_i\|$ centered in c_i.

Observe that the points $p_1, p_2, \ldots, p_i$, where $p_1 = c_1$, are not necessary distinct. And the point c_i is equal to the average of the points $p_1, \ldots, p_i$: $c_i = \sum_{j=1}^{i} p_j \big/ i$.

Proposition 2.1 ([3]). *Let c^* and R^* be the center and the radius of the smallest ball enclosing all the points of K. Then $\|c_i - c^*\| \leq R^*/\sqrt{i}$.*

Proposition 2.1 and the triangle inequality imply that the ball of radius $R(c_i, K)$ centered in c_i is a $(1 + 1/\sqrt{i}\,)$-approximation for the problem SEB. In fact, the above algorithm is a fully polynomial-time approximation scheme (FPTAS) that computes a $(1 + \varepsilon)$-approximate solution of SEB in $O(nd/\varepsilon^2)$ time for any $\varepsilon > 0$.

Describe an algorithm for the original problem Sk-EB. Let c^* and R^* be the center and the radius of the smallest ball enclosing at least k of the points. Clearly, this ball is an optimal solution of the problem SEB on the k-point set K^* the ball covers. Then by Proposition 2.1, the average of some points $p_1, \ldots, p_i \in K^*$ is at distance $R^*/\sqrt{i}$ from the point c^*.

Initially, we have no any information about the points $p_1, \ldots, p_i$ (and about the whole set K^*), but we know that these points are in the set X. The idea of the algorithm is brute-force searching for all the sequences of length i in the set X to find that sequence $p_1, \ldots, p_i$, whose average is close to c^*.

Algorithm for Sk-EB Let i be an arbitrary positive integer and $t^i : \{1, \ldots, n^i\} \to X^i$ be a enumeration of all the sequences of length i in the set X.
`Step` s, $s = 1, \ldots, n^i$: Consider the sequence $t^i(s)$, say $t^i(s) = p_1, \ldots, p_i$. Define a point $c^i(s) = \sum_{j=1}^{i} p_j \big/ i$, find a set $K^i(s)$ of the k points of X nearest to $c^i(s)$, and obtain the radius $R(c^i(s), K^i(s)) = \max_{x \in K^i(s)} \|x - c^i(s)\|$.
`Output`: The ball of radius R^i centered in c^i corresponding to the minimal radius $R(c^i(s), K^i(s))$ obtained at steps $s = 1, \ldots, n^i$.

Theorem 2.2. *Let R^* be the radius of the smallest ball enclosing k of the points of X. Then $R^i/R^* \le 1 + 1/\sqrt{i}$.*

Proof. As mentioned above, an optimal solution of the problem Sk-EB is also optimal for the problem SEB on the k-point set K^* this solution covers. Suppose that points $p_1, \ldots, p_i$ are chosen at steps $1, \ldots, i$ of the algorithm for SEB on the set K^*, $c_i = \sum_{j=1}^{i} p_j \big/ i$ and $R(c_i, K^*) = \max_{x \in K^*} \|x - c_i\|$. By Proposition 2.1 and the triangle inequality, it follows that $R(c_i, K^*)/R^* \le 1 + 1/\sqrt{i}$.

On the other hand, the sequence $p_1, \ldots, p_i$ is equal to some sequence $t^i(s)$ in the algorithm for Sk-EB. Then $c^i(s) = c_i$ and $R(c^i(s), K^i(s)) = R(c_i, K^i(s)) \le R(c_i, K^*)$ since the set $K^i(s)$ consists of the nearest points to c_i. Therefore, $R^i \le R(c_i, K^*)$ and we have $R^i/R^* \le 1 + 1/\sqrt{i}$. □

Estimate the running time of the algorithm. Since a choice of the k points of X nearest to $c^i(s)$ takes at most $O(n)$ operations (*e.g.* using the algorithm for the kth smallest number from n [11]), and all the operations over d-dimensional points take a time $O(d)$, the running time of the algorithm for Sk-EB is bounded by $O(n^{i+1}d)$.

Observe that for any $\varepsilon > 0$, we can take the parameter $i = 1/\varepsilon^2$ to compute a $(1 + \varepsilon)$-approximation for Sk-EB. In this case the running time is bounded by $O(n^{1/\varepsilon^2+1}d)$. Thus, this algorithm is actually a polynomial-time approximation scheme (PTAS) for the Smallest k-Enclosing Ball problem.

References

[1] P. K. Agarwal, S. Har-Peled and K. R. Varadarajan, *Geometric approximation via coresets*, Combinatorial and Computational Geometry, MSRI **52** (2005), 1–30.

[2] A. Aggarwal, H. Imai, N. Katoh and S. Suri, *Finding k points with minimum diameter and related problems*, J. Algorithms **12** (1991), 38–56.

[3] M. Badoiu and K. L. Clarkson, *Smaller core-sets for balls* In: "Proc. 14th ACM-SIAM Symposium on Discrete Alg.", 203, 801–802.

[4] D. Eppstein and J. Erickson, *Iterated nearest neighbors and finding minimal polytopes*, Disc. Comp. Geom. **11** (1994), 321–350.

[5] M. R. Garey and D. S. Johnson, *"Strong" NP-completeness results: motivation, examples, and implications*, J. ACM. **25** (3) (1978), 499–508.

[6] S. Har-Peled and S. Mazumdar, *Fast algorithms for computing the smallest k-enclosing disc* Algorithmica **41** (3) (2005), 147–157.

[7] A. V. Kel'manov and A. V. Pyatkin, *NP-completeness of some problems of choosing a vector subset*, J. Applied and Industrial Math. **5** (3) (2011), 352–357.

[8] C. H. Papadimitriou, "Computational Complexity", New-York: Addison-Wesley, 523 pages, 1994.

[9] J. J. Sylvester: *A Question in the Geometry of Situation*, Quart. J. Math. 1:79, 1857.

[10] V. Vazirani Approximation Algorithms. *Berlin: Springer-Verlag*, page 71, 2001.

[11] H. Wirth, "Algorithms + Data Structures = Programs", New Jersey: Prentice Hall, 366 pages, 1976.